e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

VOICE TO SIGN LANGUAGE CONVERSION USING MACHINE LEARNING

Sagar R*1

*1Student, Department Of MCA, PESITM, Shimoga, Karnataka, India.

ABSTRACT

This research presents a novel system that uses cutting-edge technologies to translate spoken English into sign language, facilitating communication for the people with hard of hearing and deaf. Using the PyAudio library, the system records speech and uses the Google Speech-to-Text API to convert it to text. Text analysis is accomplished by utilizing natural language processing algorithms. After processing, this text is compared to a video library of Handspeak sign language motions. The user sees a logical sign language interpretation of the spoken input created by concatenating the matched sign language videos in order. System can be used in offline and online modes, with the online version offering higher accuracy because it has access to large online video lexicon. The study emphasizes possibility of Future developments include the creation of a sign-to-speech conversion system and improved algorithm line CNN for increased accuracy, with goal of producing a reliable communication tool to the deaf and speech-impaired community.

Keywords: R-CNN, NLP, Mozilla Deepspeech, Google Cloud Voice-To-Text API, Speech Recognition.

I. INTRODUCTION

The challenges that the deaf and hard-of-hearing people encounter are substantial in a world where communication is essential. Tools for effective communication are crucial because hearing impairments affect approximately 1.3 million people in India alone. Conventional approaches, such the scarcity of sign language interpreters, are inadequate and mostly restricted to large cities. Because to close this gap, an automated system that converts translating voice into sign language is proposed in this work. By leveraging technologies such as NLP and the Google Speech-to-Text API, the system records spoken words, translates them into text, and then pulls relevant films from an extensive video library for sign language interpretation. In this method not only improve accessibility but also empowers the deaf community by offering a reliable and efficient means of exchange of ideas. The development of this system is a major step toward encouraging inclusivity for the deaf and hard-of-hearing communities and bridging the communication gap.

The paper [1] suggests a mechanism utilizing motion capture technology to comprehend human gestures and translate sign language to speech. Motion capture technology uses Microsoft Kinect to enable a person with speech impairment to make motions that the system translates into spoken words. As a result, the person with speech impairment can interact with people in an efficient manner. With a sample of 100 motions, the system achieves a 90% identification rate, demonstrating its excellent efficiency and accuracy. The paper [2] models a system that consists of four modules: a wireless communication unit, a processing unit, a sensor unit, and a speech storage unit. The APR9600, PIC16F877A, and flex sensor are integrated to create the system. Users put on gloves containing flex sensors, which react to hand movements. The microcontroller receives input from a suitable circuit response and uses the APR9600 to play the recorded speech.

The work [3] investigates the use of raw audio input to a Generative Adversarial Network (GAN).To produce face images of speakers. The model makes use of a deep Neural Network that is end-to-end trained from scratch. It makes no use of outside sources, such as a user photo, and instead creates faces from raw speech waveforms. The network is trained with naturally aligned audio and video inputs through a self-supervised method. On the voice train and test partitions, the generated images attain identification accuracy rates of 76.81% and 50.08%, respectively. The Study [4] is the voice-to-sign language conversion system (V2S) presented in this study is intended for deaf people living in Malaysia. By the use of this speech patterns derived from a general set of spectral parameters, System is trained for template-based recognition. The spectral parameters are kept as templates in database. Speech recognition is accomplished by comparing the input parameter set with templates that have been stored. The identified speech is then translated into sign language and presented in a video format. The recognition rate attained by the system was 80.3%.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

The Paper [5] emphasizes how crucial hand gestures are for conveying consumers' opinions. The suggested gesture recognition system recognizes some words and the English alphabet using flex sensors. It uses a synthesizer that uses a Hidden Markov Model (HMM) to turn text that has been detected into speech. Eight gestures are used for training the system, and tests revealed an average recognition rate is 87.5%.

II. METHODOLOGY

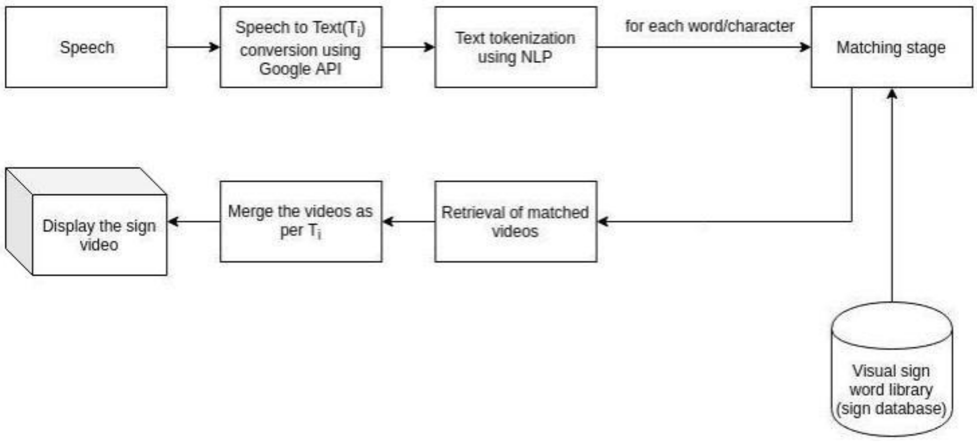


Figure 1: System Architecture

The architecture diagram represents a designed system to convert spoken language into signing language videos the process begins with capturing speech-input which is converted into text using the google speech-to-text API in order accurately match each word or character to its corresponding sign language equivalent the resulting text is processed using Natural Language Processing techniques for tokenization which breaks the text into individual words or characters. Each tokenized word or character is then compared against a visual sign word library which serves as a repository for signing language videos for each and every word the algorithm retrieves the relevant videos or characters. They are sequentially merged to create coherent sign language video finally the compiled video is displayed effectively translating into sign language the spoken content for the community of the deaf and hard of hearing this technology fills the communication gap providing a practical way to understand spoken language representation using visual sign language.

III. ALGORITHM AND TECHNOLOGY USED

- **R-CNN(Region-based Convolutional Neural Networks):**

A type of ML Algorithms called R-CNNs, or Region-Based CNN, are intended for object detection in pictures. Selective search is used by the original R-CNN to extract approximately 2,000 region proposals from a picture. After resizing each proposal, it is fed through CNN to extract features, which they are classified using an SVM and improved with linear regression for bounding boxes. Using Region of Interest (RoI) pooling to extract fixed-size feature vectors from the feature map, Fast R-CNN improves on this by processing the entire picture through a CNN to build a feature map in a single pass. Next, these vectors are categorized and refined through fully connected layers.

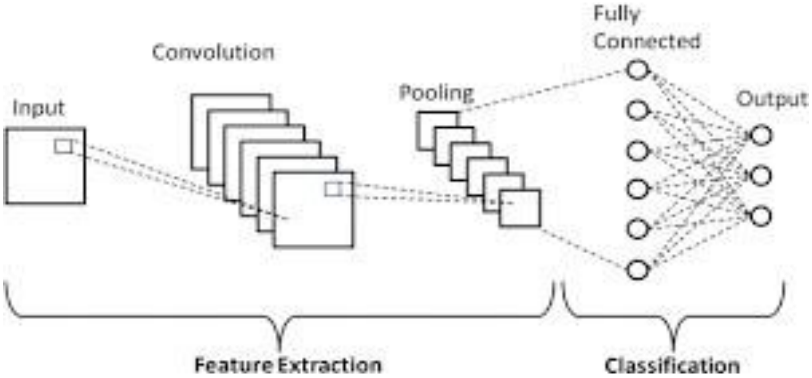
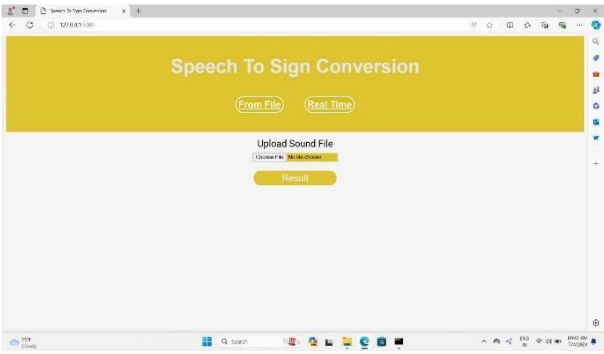


Fig 1: R-CNN architecture

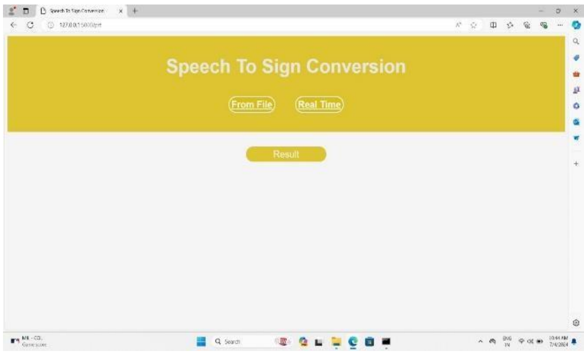
Start with converting audio input into text using Mozilla Deep Speech or Google Cloud voice-to-Text API in order to develop a system that converts voice to sign language. Installing the Python client library, turning on the Speech-to-Text API, and creating a Google Cloud account are prerequisites for using the Google Cloud API. Installing DeepSpeech and transcribing using a trained model are the two steps of Mozilla DeepSpeech. After transcribing the audio, evaluate the text and comprehend its context and semantics using NLP tools such as spaCy and NLTK. Installing and utilizing SpaCy with pre-trained language models is possible, and NLTK can help with text processing even more. Next, utilizing a database of sign language movies and pre-trained models, the processed text is mapped to signing language motions.

IV. RESULTS AND DISCUSSIONS

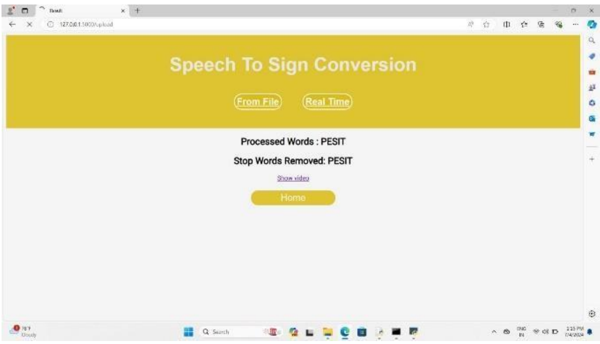
The system facilitates effective communication by converting translating spoken speech into hand sign through a seamless process. When a user speaks into a microphone, the application captures the audio and transfers it to the backend server. There, the audio is processed by a speech recognition module to convert it into text. This text is then analyzed by an NLP module to know the context, semantics, and grammar, identifying key phrases and expressions for accurate translation. The refined text is mapped to sign-gestures using a pre-trained model and a database of sign language videos. The appropriate signs are selected according to the context and semantic understanding supplied by the NLP module. Finally, the translated gestures are displayed on the user interface using video clips.



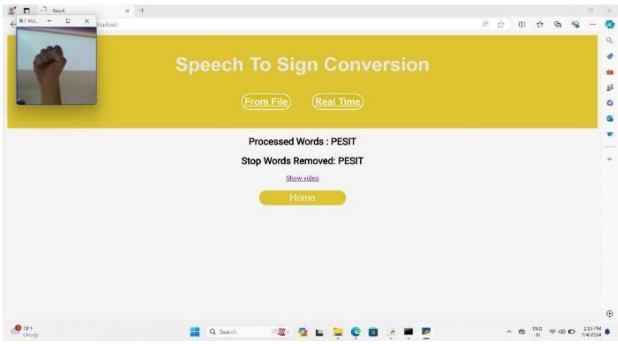
(a)



(b)



(c)



(d)

Fig: (a) and (b) represents UI for selecting Audio(speech) file input , Fig(c) Processes and recognizes the speech input , Fig(d) Shows the hand sign gesture corresponding to the speech detected.

V. CONCLUSION

Who are deaf or dumb they use signing language to express what they wish to say. It serves as a fundamental form of communication for the deaf and dumb, without which it is difficult to interact with regular people. Therefore, Creating an autonomous software system that can able to convert voice to sign language is the primary objective of this project, enabling deaf people to communicate successfully with others as well as with themselves. Using the PyAudio library, the automated system records speech, converts it to text, tokenizes it, and then uses NLP to match it with videos of sign language from the Visual Sign Word Library. It then retrieves each individual matched video and concatenates them all to display the results. When the offline and online modes are compared, it is discovered that the online mode offers superior state-of-the-art accuracy. With improved neural network or ML algorithm for data training, as well as a signing language to speech conversion platform, the approach can be expanded in future to effectively assist the deaf as well.

VI. REFERENCES

[1] Vijayalakshmi, P., and Aarthi, M. Sign language to speech conversion. In 2016 International Conference on Recent Trends in Information Technology (ICRTIT) (2016), IEEE, pp. 1–6.

[2] Foong, O. M., Low, T. J., and La, W. W. V2s: Voice to sign language trans- lation system for malaysian deaf people. In International Visual Informatics Con- ference (2009), Springer, pp. 868–876.

[3] Gaikwad, S. K., Gawali, B. W., and Yannawar, P. A review on speech recog- nition technique. International Journal of Computer Applications 10, 3 (2010), 16–24.

[4] Gunasekaran, K., and An, R. Sign language to speech Translation system using pic microcontroller.

[5] Rajaganapathy, S., Aravind, B., Keerthana, B., and Sivagami, M. Conver- sation of Sign language to Speech with human gestures. Procedia Computer Science 50 (2015), 10–15.